Scheduled to appear in 2017 in: V. Ziegler-Hill & T.K. Shackelford (Eds.), The SAGE Handbook of Personality and Individual Differences. Thousand Oaks, CA: Sage. (accepted 11/11/16)

Interpersonal Accuracy in Trait Judgments

Tera D. Letzring & David C. Funder

The SAGE Handbook of Personality and Individual Differences

INTRODUCTION

Judgments about traits and states can have important consequences. For example, whether one decides to approach a stranger at a social gathering, who one asks to watch one's children, and which job applicant is hired for a new position are all influenced by judgments of the traits and states of others. Someone who is perceived as friendly and in a good mood is more likely to be approached at a social gathering; someone who is perceived as responsible, not deceptive, and valuing security is more likely to be asked to watch one's children; and someone who is perceived as trustworthy and valuing achievement is more likely to be offered a job. Across these and other domains, research has investigated the process of making judgments, the degree of accuracy of such judgments, and a host of variables related to accuracy. In the present chapter, following the recommendation of Hall, Schmid Mast, and West (2016), the accuracy of trait and state judgments will collectively be referred to as *interpersonal accuracy*. Past research has referred to the people making the judgments as *judges*, *perceivers*, and *decoders*, while the people being judged have been referred to as *targets* and *encoders*. In this chapter, the terms judges and targets will be used.

Person perception is a broad area of research that investigates both the processes that take place when one person perceives another and the bias and accuracy of the judgments that result. Ordinary observers can perceive many aspects of other people, such as dispositions and characteristics that are relatively stable across situations and over time, including broad personality traits such as the Big Five personality dimensions (i.e., extraversion, agreeableness, conscientiousness, neuroticism/emotional stability, and openness to experience/intellect; Biesanz, West, & Millevoi, 2007; Funder, 2012; Letzring & Human, 2014) or more specific aspects of people such as trustworthiness (Rule, Krendl, Ivcevic, & Ambady, 2013), personal values, morals, or character (Goodwin, Piazza, & Rozin, 2014; McDonald & Letzring, in press), and how risky people are and how they perceive risk levels (Mishra & Sritharan, 2012; Vineyard, 2016). People also form perceptions of psychological states that change

relatively frequently (compared to traits), including emotion or affect (Banziger, 2016; Hall, Andrzejewski, Murphy, Schmid Mast, & Feinstein, 2008), thoughts and feelings (Hodges, Lewis, & Ickes, 2015; Ickes, 2016), and deceptiveness vs. truth-telling (Burgoon & Dunbar, 2016). Person perception is so broad a topic that in order to provide a meaningful explanation and discussion, as well as to maintain a reasonable length, this chapter focuses on examining judgments of personality traits. However, it should be noted that many of the general points made in this chapter, especially explications of processes and analyses, also apply to other kinds of judgments.

Personality traits are characteristic patterns of cognitions, emotions, and behaviors that are relatively stable over time, such that the states people typically experience reflect and can be predicted from their trait level within a domain (Barenbaum & Winter, 2008; Fleeson & Jayawickreme, 2015; Fleeson & Law, 2015). Ample evidence shows that both informal judgments and formal assessments of traits can predict a variety of outcomes up to several decades in the future, including individual outcomes such as happiness and health, interpersonal outcomes such as romantic and non-romantic relationship satisfaction, and institutional outcomes such as job satisfaction and criminality (Ozer & Benet-Martinez, 2006). Moreover, personality traits can predict important life outcomes such as divorce and mortality about as well as these same outcomes can be predicted by socioeconomic status and cognitive ability (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Because traits have such important implications, it is only to be expected that the ability to accurately judge them would be related to many positive characteristics and outcomes.

In this chapter, we will provide an overview of the history of research on the interpersonal accuracy of trait judgments, explain the ways in which accuracy has been assessed, and describe the most prominent conceptual models used to understand judgments of traits. Then we will describe empirical findings regarding cues that are commonly used to make accurate judgments, discuss the role of stereotypes and projection in judgments, briefly discuss meta-accuracy, outline several characteristics and constructs that correlate with accuracy, and summarize some methods for improving accuracy and the efficacy of those methods. Finally, we will end with a summary and suggestions for future research directions.

FROM ACCURACY TO BIAS AND BACK

Research on interpersonal accuracy for personality traits has a rocky history: it was highly active in the early-1900s, was harshly criticized and fell away to almost nothing in the mid-1900s, started gaining ground again in the 1980s, and had returned to a respected and active field of study by the early 2000s.

Early work on person perception accuracy

Research on person perception accuracy was first published in the early 1900s. Adams (1927) focused on the attributes of the *good judge*, or in other words, people who consistently make relatively accurate personality judgments. Adams found that good judges of the self had several beneficial attributes such as mental ability and social skills, but that good judges of others had both positive and negative attributes, such as having high mental ability while also being touchy and quick-tempered. Vernon (1933) focused on the attributes of the good judge of the self, friends, and strangers (based on photographs), and found that many factors were associated with individual differences in accuracy. Vernon concluded that a general trait of judgmental ability did not exist. At the same time, however, he also found that some attributes were consistently related to accuracy, including intelligence and being artistic. Allport (1937) and Taft (1955) agreed with Vernon that interpersonal accuracy is more specific than general, but likewise identified several characteristics related to accuracy, again including intelligence and artistic interests, and also social skills.

Allport (1937) identified several further issues relevant to judgments of personality, including the moderators of accuracy, the automaticity of first impressions, and the importance of nonverbal cues in making judgments of others. Around the same time, Estes (1938) identified three factors that influence accuracy of judgments: the ability of judges, the variability in how easily targets could be judged, and the characteristic being judged. Allport (1937) identified judgmental ability as an individual difference and Estes presented results that supported this finding, and this notion is largely retained today. Work of this type foreshadowed later conceptualizations of the process or stages of making accurate judgments (see section on the Realistic Accuracy Model).

Taft (1955) discussed projection and stereotypes as confounds of self-other agreement, and may have laid the initial groundwork for the work presented later that same year by Cronbach (1955). Both Taft and Cronbach identified multiple components of accuracy scores, not all of which are what accuracy researchers intended to study. Even with this overlap in the content of these articles, the movement away from studying accuracy of trait judgments is typically identified as being instigated by Cronbach (Hall & Colvin, 2011).

The movement away from studying accuracy and real people

In many histories of research on the accuracy of personality trait judgments, the claim is made that research devoted to this topic fell away shortly after two articles were published in 1955 (i.e., Cronbach, 1955; Gage &

Cronbach, 1955). However, the decline in research was not as complete or as devastating as is sometimes described. Hall and Colvin (2011) summarize convincing evidence that research on person perception across many domains, including personality traits, was conducted by prominent researchers in the field and published in respected journals during this time(Hjelle, 1968). However, even if there was not a complete exodus away from research on accuracy of personality judgment, the topic was for a time neither prominent nor prestigious. The leading textbook on person perception of that era explicitly stated that

> The accuracy issue has all but faded from view in recent years, at least for personality judgments. There is not much present interest in questions about whether people are accurate…. On the other hand, in recent years there has been a renewed interest in how, why and in what circumstances people are inaccurate. (Schneider, Hastorf, & Ellsworth, 1979, p. 224)

Instead of studying accuracy, researchers moved to areas that did not require them to calculate accuracy scores, thereby avoiding the problems presented by Cronbach and Gage. This movement contributed to a greater focus on research in two areas: examinations of the process of making judgments of others and the assembly of a vast catalog of errors and biases in judgments (Funder, 1987).

Researchers who studied the cognitive processes that occur during interpersonal judgment were not concerned with accuracy and therefore had no need to calculate accuracy scores (Hastie & Rasinski, 1988). Heider (1958) and Kelley (1967, 1973), among others, were early proponents of attribution theory, which focuses on how people explain events, especially for behaviors of the self and others, and whether these are best explained in terms of something about the person or something about the situation (Kelley, 1973). Research on attribution theory examined the circumstances - in terms of time, persons, and entity - that typically led to different types of causal judgments. The issue of whether, when, or under what circumstances these judgments are right or wrong was completely bypassed.

Other descriptions of the process of judging the causes of the behaviors of other were proposed by Trope (1986, 1989) and Gilbert (1988). Trope's model includes two stages: 1) identification of what behavior has occurred and 2) determination of why the behavior was performed. Gilbert's model includes three stages: 1) identification of the intent of a behavior, 2) attributions for the behavior, primarily in terms of dispositional inferences or personality judgments, and 3) an integrated impression of the target based on dispositional inferences from the previous stage and corrections based on situational information. Both models were tested using fictitious behaviors, situations, and

targets as the judgmental stimuli. For example, Trope (1986) presented judges with faces that either showed ambiguous or unambiguous emotions paired with descriptions of emotion-arousing situations, and found that emotion judgments were consistent with situational descriptions, even when different descriptions were given for the same emotional expression. Judges were also asked to identify traits of the targets based on the facial expressions and situational descriptions, and it was found that trait judgments were influenced by affect cues. However, as was typical in this research tradition, accuracy of the trait judgments was not examined.

Another common area of research on person perception during this time focused on biases and errors in judgments (Funder, 1987; Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980). One of the most well-known errors has been called the *fundamental attribution error* or *correspondence bias* (Ross, 1977; Ross, Amabile, & Steinmetz, 1977). The main tenet of this error was that, when explaining behavior, people rely too much on dispositional causes and not enough on situational causes. However, what was meant by "too much" or "not enough" was seldom specified; often the implicit assumption underlying the research was that any dispositional attribution could be assumed to be wrong. To our knowledge, no research on the fundamental attribution error ever directly compared the actual relationship of personality to a behavior with the perceived relationship of personality to that same behavior, in order to show that the latter is greater than the former. Therefore, research on the fundamental attribution error did not illuminate whether the judgments it examined were accurate or inaccurate (Harvey, Town, & Yarkin, 1981). In related work that also addressed judgmental processes rather than accuracy per se, Tversky and Kahneman (1973) became well-known for their work on heuristics involved in decisions about probabilities or frequencies. While these heuristics were often viewed as sources of bias, in fact Tversky and Kahneman explicitly noted that heuristics can and often do promote accurate as well as inaccurate judgment (Gigerenzer & Gaissmaier, 2011).

The return of the study of accuracy

Even while research on error and biases continued to gain momentum, some researchers advocated for a return to the study of accuracy (Funder, 1980, 1987; Hastie & Rasinski, 1988). In particular, researchers renewed the search for evidence that could assess whether and when ratings of personality provided by lay persons could provide valid and useful information (Funder, 1983; Kammann, Smith, Martin, & McQueen, 1984). Furthermore, some researchers had begun to realize that learning about biases and errors did not necessarily increase knowledge about how to make accurate judgments, because a process that leads to an error in the laboratory could lead to an accurate

judgment in the real world (Funder, 1987, 1995). Beginning in the late 1980s and early 1990s, and accelerating during the early 2000s, an increasing number of researchers heeded the call for an increased understanding about how accurate judgments could be made.

ASSESSING ACCURACY

Accuracy can be assessed in multiple ways, and the specific method that is used has a critical impact on the meaning of the findings because the conceptual definition of accuracy depends upon the method by which it is assessed. Factors to consider are the criterion that is used to establish accuracy, whether accuracy is computed for a single item across a set of targets or judge-target pairs vs. for a set of items for judgments of a particular target, and the analytic strategy that is used to estimate or compute accuracy.

Consider, first, the accuracy criterion. The simplest criterion, and therefore the one that is most frequently used, is the self-rating of the target. When ratings from judges are compared to this self-rating, the outcome is referred to as *self-other agreement*. However, self-ratings can be biased and the typical way to decrease the bias and increase the reliability of the accuracy criterion is to average the self-ratings with ratings provided by *acquaintances* or *informants* of the targets. Ideally, the acquaintances will have known the target for a long enough time to be familiar with him or her and able to provide accurate ratings of his or her personality. Some studies set a lower limit for the length of acquaintanceship, such as six months (Letzring & Human, 2014), whereas other studies ask targets to identify people who know them well and would be willing to provide ratings of their personality (Funder & Colvin, 1988; Letzring, Wells, & Funder, 2006). Ratings from clinical psychologists following an hour-long life-history interview have been included in accuracy criteria (Letzring et al., 2006). Behavioral data can also be used as a criterion (Vazire, 2010), which is especially useful for examining the accuracy of self-judgments and for comparing accuracy of self-judgments to accuracy of other-judgments. When the criterion is composed of a composite of ratings, the resulting accuracy score is typically referred to as *realistic accuracy* to denote that the criterion is closer to the targets' real or true personality than is possible with a self-report or any other single indicator (Funder, 1995; Letzring et al., 2006).

It is also possible to estimate accuracy by correlating judgments of two or more people of the same target person, which is referred to as *consensus* or *peer-peer agreement* (Funder & Dobroth, 1987; Kenny, 1991, 1994; Kenny, Albright, Malloy, & Kashy, 1994). However, judgments that reach high levels of consensus are not necessarily accurate because it is possible – at least in principle – for a set of judges to agree with each other and all

be inaccurate. Furthermore, self-other agreement and consensus sometimes show important differences in relations to other factors. For example, judges who watched longer recordings of targets showed increases in accuracy, but not consensus, compared to judges who watched shorter recordings; and judges who had known the targets for months or years achieved higher accuracy than judges who had watched 5-minute videos of the targets, but consensus was more similar between these levels of acquaintance (Blackman & Funder, 1998; Funder & Colvin, 1988). For this reason, in this chapter we focus on research that has examined self-other agreement or realistic accuracy, but not consensus.

Regardless of the accuracy criterion that is used, accuracy scores can be computed using a variable-centered or person-centered approach (Back & Nestler, 2016). Item-level correlations are used for a variable-centered approach in which the aim is to assess how accurately single items or traits are rated, or in other words, how accurately a judge can order a set of targets for a single item or trait. Item-level accuracy scores can be calculated for a single judge across a set of targets, which results in an accuracy score that reflects the level of consistency between how the judge ordered the set of targets and the ordering of those targets based on the accuracy criterion. The sample size for the correlation is the number of targets. An advantage of this approach is that the resulting score is associated with a single judge, and it also avoids the problems associated with profile correlations, most notably, stereotype accuracy (Cronbach, 1955; Gage & Cronbach, 1955). Furthermore, it is possible to average the correlations for several items to create an accuracy score for a trait. A disadvantage of this approach is that a large number of targets are required in order for the resulting correlation to be stable (Schönbrodt & Perugini, 2013), which will place a heavy burden on judges if they are observing recordings of many targets and/or judging a large set of items for each target.

Item-level accuracy can also be computed across pairs of judges and targets, which may be necessary when judges each rate one or only a few targets. This correlation reflects the consistency between the ordering of a set of judgments and the criteria for the corresponding set of targets. The sample size for each correlation is the number of judge-target pairs. An advantage of this approach is that a stable accuracy score can be computed for a single item, as long as there are a sufficient number of judge-target pairs. A disadvantage of this approach is that it does not yield an accuracy score that can be associated with a particular judge (Funder, 1999).

Profile correlations are used with a person-centered approach in which the aim is to assess the level of consistency between how a set of items or traits is ordered for a single target and the ordering of those same items or

traits for that same target according to the accuracy criterion (Furr, 2008). Large profile correlations indicate that the two profiles have similar shapes. However, profiles that are very similar to the profile of the average person (i.e., profiles that are highly normative or stereotypical) are likely to result in inflated correlations (Furr, 2008). While this inflation is sometimes considered to be an artifact, it is reasonable to assume that an important and valid part of the knowledge necessary to accurately judge an individual is an understanding of what people in general are like.

The typical method for dealing with the influence of normativeness is to subtract the normative profile from the accuracy criteria, which results in a distinctive profile that reflects how the target differs from the average person. Judgments can then be correlated with this distinctive profile to reflect the level of consistency between how a set of items or traits is ordered for a single target and the ordering of those same items or traits for that same target in terms of differences from the average person, which is referred to as *distinctive accuracy*. Additionally, judgments can be correlated with the normative profile itself to determine the level of consistency between judgments and the normative profile, which is referred to as *normative accuracy*.

Again, while it is sometimes useful to analyze distinctive and normative accuracy separately, normative accuracy is more than simply an artifact and an exclusive emphasis on distinctive accuracy can be misleading (Funder & Colvin, 1997). For example, a judge can attain a high distinctive accuracy score only by making accurate judgments of the unusual aspects of unusual people; making accurate judgments about the normative aspects of normal people does not increase one's distinctive accuracy score. Yet, by definition, most targets are statistically "normal," so distinctive accuracy only captures part – and perhaps only a small part – of what is ordinarily meant by accurate judgment in everyday settings. Furthermore, the normative profile is highly favorable, so judges could also attain high levels of normative accuracy by judging others in a favorable manner (Wood & Furr, 2016).

Research designs sometimes ask each judge to rate several targets, a practice which can increase the reliability of the accuracy scores, increase the generalizability of the findings, and make it possible to separate systematic differences in accuracy between judges from judge-target interactions and error variance. However, this methodology introduces dependencies in the data when computing item-level correlations using judge-target pairs or profile correlation, due to the same judges rating multiple targets. Furthermore, some studies use a standard set of targets that are rated by all judges, which introduces even more dependencies into the data. Round-robin designs, in which a group of participants all rate themselves and each other, are also sometimes used in accuracy research. Kenny and colleagues developed the Social Relations Model (SRM) to analyze this type of data structure (Kashy &

Kenny, 2000; Kenny & Kashy, 2014; Kenny & la Voie, 1985; Warner, Kenny, & Stoto, 1979). SRM is essentially a modified two-way random effects analysis of variance that partitions the total variance into three parts: perceiver, target, and relationship. The *perceiver effect* is how a perceiver, or judge, typically judges all targets in terms of their standing on an item or trait. The *target effect* is how a target is typically judged, again in terms of their standing on an item or trait. The target effect can be used to estimate consensus. The *relationship effect* is how a specific judge views a specific target, or the uniqueness of a judge's ratings of a target. SRM analyses have been used by many researchers to examine dyadic and round-robin data. For example, SRM was used to examine perceptions from studies that used a speed dating format in which people interacted in several dyads and rated their partner after each interaction (Finkel & Eastwick, 2008), as well as to examine how well trait ratings and behavioral predictions about extraversion made without any interaction would predict behavior during subsequent interactions (Levesque & Kenny, 1993). Significant levels of relationship variance were found for ratings of extraversion, agreeableness, conscientiousness, emotional stability, and culture (Levesque & Kenny, 1993).

Another way to analyze data with a hierarchical structure is to use a multilevel modeling (MLM) framework. The Social Accuracy Model (SAM; Biesanz, 2010)was designed for this purpose, and integrates SRM and Cronbach's approach for dealing with the components of accuracy (also see Biesanz, this volume). SAM decomposes ratings of several targets on several items into effects due to the judges, which is called *perceptive accuracy*, and effects due to the targets, which is called *expressive accuracy*. Judges with high perceptive accuracy tend to judge others accurately, and targets with high expressive accuracy tend to be judged accurately by others. Each of these types of accuracy can be divided into *normative accuracy* and *distinctive accuracy*. Normative accuracy reflects the similarity between judgments of individual targets and the average or normative profile. The normative profile is typically computed by averaging the accuracy criteria for each item across a large set of targets, usually many more targets than were rated by each judge, to ensure a stable average. Distinctive accuracy reflects the similarity between judgments of a target and how the accuracy criterion for that target differs from the normative profile. These are all types of profile accuracy. A SAM analysis is accomplished by predicting the judges' ratings from the normative profile and the distinctive profile (after subtracting out the normative profile). The hierarchical structure is used to account for judges rating several targets and targets being rated by many judges. Characteristics of judges or situations can be added to the multilevel equation to examine how such variables moderate normative and distinctive accuracy. A moderator with a significant interaction with the normative profile is evidence that

normative accuracy differs across levels of the moderator, and a moderator with a significant interaction with the accuracy criterion profile is evidence that distinctive accuracy differs across levels of the moderator. For example, the positive interaction between perceptive normative accuracy and judges' agreeableness is evidence that more agreeable judges achieve higher levels of normative accuracy than less agreeable judges (Letzring, 2015), and the interaction between expressive distinctive accuracy and targets' adjustment is evidence that better-adjusted targets are perceived with higher levels of distinctive accuracy than targets with lower levels of adjustment (Human, Biesanz, Finseth, Pierce, & Le, 2014).

Yet another approach for analyzing accuracy data is through the use of cross-classified structural equation modeling (CC-SEM; Nestler & Back, 2015). This method follows a trait-based or item-level approach, which differs from the person-based or profile approach used in SAM. SEM deals with the hierarchical or clustered nature of the data, and the cross-classified nature of the model is used to deal with the structure of the data in which targets are clustered within judges, while judges are also clustered within targets. Using a CC-SEM approach, the criterion and the predictors are modeled as latent variables. Furthermore, a Bayesian approach of estimation "allows for random intercepts across clusters, random slopes across clusters, unequal cross-cluster sizes, and missing data" (Nestler & Back, 2015, p. 9). This approach can be used to examine how characteristics of judges moderate levels of accuracy, as well as how the availability of cues mediates accuracy. A primary difference between the CC-SEM and SAM is that CC-SEM does not decompose accuracy into distinctive and normative components, whereas SAM does.

CONCEPTUAL MODELS OF ACCURATE JUDGMENTS

Several models of the process of accurate judgment have been proposed over the years, with revised and new models still appearing. These models can be used to organize research questions and findings regarding the process of how accurate judgments are made and the moderators of accuracy.

Brunswik's lens model and the dual lens model

One longstanding model used for studying accuracy of person perception is the lens model that was developed by Brunswik (1955, 1956). In this model, multiple cues are given off by an object – in the case of person perception, the object is the target or a particular characteristic of the target. Cues differ in their level of relevance to the object, and *cue validity* or *ecological validity*[1] refers to the degree to which a cue is relevant to the object. For example, showing up to an appointment on time would be a valid cue for conscientiousness, but probably not a valid cue for openness. These cues can be used by judges to arrive at a judgment of the object or trait of interest. *Cue*

utilization refers to how strongly each cue relates to the final judgment. It is important to keep in mind that research on the lens model does not ask judges which cues they are using, but instead infers this use through examining correlations between cues and judgments. The term *achievement* describes judgments that are consistent with the object. Achievement, or accuracy, will be high when many valid cues are available and the judge utilizes valid cues and does not utilize invalid cues.

This model has been used by many researchers over several decades to examine achievement, cue validity, and cue utilization for various characteristics, including estimates of IQ scores of clinical patients based on responses to the Rorschach (Hammond, 1955; Hammond, Hursch, & Todd, 1964), the Big Five traits and intelligence based on observing targets walk to a table, read a weather forecast, and walk out of the room (Borkenau & Liebler, 1992, 1995), deception based on cues such as vocal pitch, rate of speech, and appearing incompetent (Hartwig & Bond, 2014), and the Big Five traits based on a still photograph, a still photograph plus a trait-descriptive sentence, and a short video (Beer & Watson, 2010). For example, Beer and Watson (2010) found that cues were most effectively utilized when judges watched a video of targets, as compared to viewing a still photograph or a still photograph plus a trait-descriptive sentence, and valid cues that were utilized for judgments of neuroticism, extraversion, and agreeableness in the video condition were fashionable dress and attractiveness. The lens model has been used to examine how situational factors and characteristics of tasks influence judgments. Based on a meta-analysis, average accuracy scores across studies are greater than zero; accuracy is lower when there are more cues available to judges, accuracy is lower when judges have to infer cues vs. when they are given cues, and accuracy is lower when judges have higher initial levels of expertise (Karelaia & Hogarth, 2008).

An expansion of Brunswik's lens model is the dual lens model (DLM; Hirschmuller, Egloff, Nestler, & Back, 2013; Nestler & Back, 2013). This model separates the cue validity part of the model into explicit and implicit aspects by identifying cues as either controlled or automatic. The cue utilization part of the model is likewise separated into deliberate and intuitive judgments. This decomposition of the parts of the lens model allows additional questions to be addressed, such as what are the levels of self-other agreement for explicit and implicit trait levels, how consistently do judges use controlled and automatic cues to make judgments of explicit and implicit trait levels, and what factors moderate self-other agreement for explicit and implicit trait levels? An initial test of the model found that judges could achieve statistically significant levels of self-other agreement for both explicit and implicit measures of extraversion when decisions are both deliberate and intuitive, controlled cues were more

strongly associated with judgments of explicit extraversion while automatic cues were more strongly associated with judgments of implicit extraversion, and an intuitive vs. deliberate judgment style moderated some aspects of self-other agreement (Hirschmuller et al., 2013).

The Weighted Average and PERSON models

Kenny (1991) proposed a model of consensus of personality judgment that he called the Weighted Average Model (WAM)(WAM; Kenny, 1991), which he later updated to a model called PERSON (Kenny, 2004). *Consensus* is the amount of agreement between two or more judges about a target or set of targets (see section on Assessing Accuracy for a more detailed discussion of consensus). The original name – WAM – was based on the proposition that people make trait judgments based on the scale values and weights they assign to observed behaviors of targets. Kenny outlined six parameters that influence levels of consensus: 1) the amount of information or number of behaviors observed by the judges (which is referred to as *acquaintance*), 2) the amount of *overlap* in the information on which judges base their ratings, 3) the *similarity* of the judges' meaning systems, 4) the *consistency* of the target's behavior across situations, 5) *extraneous information* that is not based on the target's behavior but is used in the judgment, and 6) the amount of *communication* between judges in which they discuss their impressions of the target(Kenny, 1991). Additionally, each judge has a unique impression of the target that is not based on the behaviors of the target, and this also affects consensus. Consensus will be high when judges assign the same values to observed behaviors and when judges have access to mostly the same information.

WAM can also be applied to accuracy of judgments. Consensus is necessary for accuracy, but is not sufficient for demonstrating accuracy. In other words, in order for two judges to both have accurate perceptions of a target, they must have a high level of consensus; but two judges who agree with each other are not necessarily accurate, as both of the judges could be wrong. According to Kenny (1991), in order for consensus to be used to determine accuracy, judges must not communicate with each other about their impressions of the target and judges must be exposed to entirely different behaviors of the target. Furthermore, as judges observe more behavioral acts of a target that are consistent across situations and if there is perfect overlap in the behaviors that are observed by judges, accuracy could increase while consensus stays the same. Researchers have examined the influence of some of the sources of variance identified in WAM on accuracy. For example, Beer (2013) examined accuracy of judgments based on photographs for groups of 2, 3, and 4 judges who were allowed to communicate in order to

arrive at consensual judgments for each target. Ratings based on communication among judges were sometimes more accurate than ratings provided by a single judge, but not for all traits and all group sizes.

The acronym PERSON stands for personality, error, residual, stereotype, opinion, and norm, which are proposed to be the sources of variance in person perception (Kenny, 2004). *Personality* is the general way in which a target is perceived and is hypothetical because it is based on all judges observing all of the target's behavior; *error* is the "part of the judgment of behavior that is not shared with other perceivers and not correlated with the judgment of other target behaviors" (Kenny, 2004, p. 268); *residual* is the judges' non-shared stereotypes about physical appearance; *stereotype* is the judges' shared assumptions about physical appearance; *opinion* is the judges' unique perception of the target; and *norm* is the unique meanings that each judge assigns to behaviors and "reflects inconsistencies or changes in the target's behavior" (p. 269).

Kenny (2004) also identified two main sources of information about others: categorical, which is based on appearance and demographics; and behavioral, which is based on observed behavioral acts. Kenny proposed that these sources of information are integrated with each other to make a final judgment. As judges are exposed to more behavioral information, they place less weight on categorical information and more on behavioral information. This view is consistent with research on the role of stereotypes in judgments of others (see section on The Role of Stereotypes and Projection), although Kenny was careful to distinguish between categorical information and stereotypes, with categorical information always being nonverbal and learned at the first exposure to a target.

Both WAM and PERSON lead to the proposal that acquaintance, or the length of time a judge has known a target, has less influence on accuracy than is typically expected (Kenny, 2004; Kenny et al., 1994). In particular, increases in acquaintance are useful very early in a relationship, but accuracy does not continue to increase over longer stretches of time. This proposition has been empirically supported by calculating consensus within 32 studies as the ratio of variance due to the target to total variance (Kenny et al., 1994, but see the discussion of "good information" and the acquaintanceship effect in the summary of the Realistic Accuracy model, later in this chapter). A more recent study examined how providing judges with additional information about targets affects accuracy, and found that providing a descriptive sentence of the targets' level of agreeableness significantly increased accuracy of judgments of agreeableness in relation to judgments based on only a photograph, while accuracy for neuroticism and conscientiousness also increased but not to a statistically significant degree (Beer & Watson, 2010). However,

providing video information did not significantly increase accuracy for judgments of extraversion and agreeableness in relation to judgments based on only a photograph.

Kenny (1991, 2004) acknowledged some limitations of both the WAM and PERSON models, in particular the existence of several simplifying assumptions that are unlikely to be true in real judgments, such as all acts having equal weights. He claimed that such assumptions are necessary to understand the complex phenomenon of person perception. Kenny also pointed out advantages of these models, which include acknowledging that perceptions of others include both accurate and inaccurate components, and that these models grapple with the complexity of considering multiple acts or behaviors when forming a judgment, which is more consistent with reality than models that focus on a single act.

The Realistic Accuracy Model

The Realistic Accuracy Model (RAM) specifies the stages that must be successfully completed for an accurate judgment of personality traits to be possible, along with four factors that moderate levels of accuracy (Funder, 1995, 2012). The stages are relevance, availability, detection, and utilization; and the moderators are the good judge, good target, good trait, and good information. "Good" is used to describe each moderator to highlight the focus on accuracy rather than error (Funder, 1995). This model is quite different from cognitive models that were popular when RAM was initially proposed. RAM focuses on the entire process of making an accurate judgment and includes processes in the social environment between targets and judges, whereas cognitive models tend to focus on the utilization stage and what judges do with cues that have been explicitly provided by the researcher. RAM has a similar limitation as WAM and PERSON, in that it is a simplified model of what happens in the real world, in this case by explaining the process as if one cue were being used to judge one trait. In reality, many cues are available to judges and many traits are judged simultaneously, often multiple cues interact and combine when utilized by judges, and judgments of one trait may influence judgments of other traits.

Stages of RAM. The first stage of RAM is *relevance*, which denotes that cues given off by the target must be relevant to the trait being judged. This stage is consistent with cue validity in Brunswik's lens model. Cues can consist of many things, including behaviors, thoughts and feelings, spoken words, facial expressions, physical appearance, clothing, etc. The second stage is *availability*: only cues that are available to the judge can be used to make an accurate judgment. Overt behaviors, spoken words, and physical appearance are available, whereas unrevealed thoughts and feelings are not. Also, only cues that are present in the settings that are shared by the judge

and target are available; for example, behaviors that are performed only at home will not be available to colleagues who know the target only in the workplace. These first two stages focus on the targets of judgment, in that the targets have to make these relevant cues available in some way, either intentionally or unintentionally. For the third stage, *detection*, the focus shifts to the judge. Judges must detect the available cues that are present; they need to be paying attention and perceptually acute. The final stage is *utilization*. In this stage judges must correctly use the cues that have been detected to make a judgment, matching valid cues to the traits with which they are associated.

The stages of RAM are assumed to be multiplicative, which has some important implications (Funder, 1995; Human et al., 2014). First, it means that if any stage is not completed, then accuracy is not possible. For example, if a judge does not detect any cues from a target, then there is nothing to utilize and accuracy will be zero. Another implication is that imperfections in the attainment of each stage limit the possible attainment of the next stage, which highlights the complexity of making accurate judgments and the difficulty and perhaps impossibility of judgmental perfection.

Moderators of accuracy. Before Funder (1995) proposed RAM and the moderators of good judges, targets, traits, and information, other researchers identified some of the same moderators. As detailed in an earlier section of this chapter (From Accuracy to Bias and Back), Adams (1927), Vernon (1933), and Allport (1937) all examined the good judge moderator of accuracy. In addition, Allport(1937) (1937) and Estes (1938) separately proposed three moderators that were essentially the good judge, good target, and good trait. Allport also identified the good information moderator through his discussion of interviews as more likely to be valuable for making accurate judgments when a variety of information is sought. Funder connected these moderators to RAM by proposing that these variables influence accuracy by modifying success at one or more of the stages of judgment.

Most work on the *good judge* has focused on the detection and utilization stages (Funder, 1995), and the most consistent finding is that positive characteristics are positively related to accuracy, and negative characteristics are negatively related to accuracy. For example, positive characteristics related to accuracy include trustfulness, courage, sense of humor, maturity, intelligence, social skills, emotional stability, warmth, empathy, agreeableness, sympathetic and considerate, and internally consistent (Adams, 1927; Allport, 1937; Colman, Letzring, & Biesanz, 2016; Kolar, 1996; Letzring, 2008; Taft, 1955; Vernon, 1933; Vogt & Colvin, 2003); whereas negative characteristics that are negatively related to accuracy include anxious, defensive, hostile, manipulative, domineering, vindictive, avoidant, and narcissistic (Kolar, 1996; Vogt & Colvin, 2003). Characteristics and behaviors of judges

can also affect the relevance and availability stages. This was demonstrated in a study in which profile accuracy of judges who observed and rated a recorded interaction was compared for different numbers of good judges in the interaction (Letzring, 2008). The rationale for this analysis was that if good judges influence the amount of relevant and available information produced by targets, then observers of interactions with more good judges should be more accurate. This pattern of accuracy was supported in two datasets.

More recent research has also examined the characteristics of good judges for two components of accuracy: distinctive and normative accuracy (see Assessing Accuracy section in this chapter for descriptions of these components). In one study, accuracy was based on ratings following observations of recorded interactions (Letzring, 2015). Correlations with accuracy differed for the components. For normative accuracy, there were several significant relations with characteristics of judges, including agreeableness, interpersonal control, interpersonal support, positive affect, and life satisfaction. On the other hand, fewer characteristics were significantly associated with distinctive accuracy than would be expected by chance.

A second moderator within RAM is the *good target*, denoting that some people are consistently judged with higher accuracy than others. This moderator is also referred to as *judgability*. Early research supported the importance of this factor and described bad targets as "contradictory and confusing" (Cline, Atzet, & Holmes, 1972, p. 386). Good targets are typically viewed positively by both themselves and their peers, and in particular as high on extraversion, agreeableness, conscientiousness, and emotional stability (Colvin, 1993). Good targets are psychologically adjusted and exhibit positive behaviors such as warmth, compassion, and social engagement (Colvin, 1993). A more recent review focused on good targets in terms of targets who were judged with high levels of distinctive accuracy, and found that judgability was related to consistency, psychological well-being, and adjustment (Human et al., 2014). Human and colleagues also identified that the mechanism of the good target worked through the relevance stage of RAM, in that the behaviors of good targets were more consistent with their personalities than the behaviors of bad targets, and this was especially true for well-adjusted targets.

A third moderator is the *good trait*, which denotes that some traits are judged more accurately than others. Several characteristics of traits have been connected with how accurately they are judged. One aspect is *visibility* or *observability*, which reflects the number of external cues that are relevant to a trait and typically made available to others. Observability has also been defined as the extent to which items that assess a trait are descriptions of external behaviors vs. internal attitudes or beliefs. Traits that are associated with more external cues and behaviors tend to be

more accurately judged, whereas traits that are associated with more internal manifestations, such as thoughts and emotions, tend to be less accurately judged (Funder & Dobroth, 1987; John & Robins, 1993; Paunonen & Kam, 2014; Watson, Hubbard, & Wiese, 2000). However, some research has found that the link between visibility and accuracy only holds with low to moderate levels of acquaintance between judges and targets but not for high levels of acquaintance (Paunonen, 1989), or that the link does not exist at all within a highly acquainted sample (McDonald & Letzring, in press). Another aspect of traits that has been examined is *ratability*, which is similar to visibility but broader in that it also includes other aspects such as whether traits describe behaviors that are likely to be performed publicly vs. privately, and the difficulty of the words used to describe a trait (Paunonen & Kam, 2014). There is evidence that ratability is positively correlated with accuracy (Paunonen & Kam, 2014; Ready, Clark, Watson, & Westerhourse, 2000).

Another aspect of the good trait is *evaluativeness* or *social desirability*. Social desirability pertains to whether having a trait is desirable, undesirable, or neutral. Desirable and undesirable traits are high in evaluativeness, and neutral traits are low in evaluativeness. Accuracy is typically higher for less evaluative traits (John & Robins, 1993), whereas accuracy is unrelated to social desirability (Paunonen & Kam, 2014). These seemingly contradictory findings could reflect a curvilinear relationship between desirability and accuracy, such that traits with moderate levels of desirability (which corresponds to low evaluativeness) are judged more accurately than traits with either high or low levels of desirability (which corresponds with high evaluativeness). This explanation has empirical support (John & Robins, 1993). However, not all research has found a correlation between evaluativeness and accuracy (Ready et al., 2000).

The combination of observability and social desirability/evaluativeness factors are commonly used to explain why some traits are typically judged with higher levels of accuracy than others. For example, extraversion is typically judged with high levels of accuracy, and it has many behavioral cues that are typically present in unstructured social interactions and physical cues that are available even in still photographs (Connelly & Ones, 2010; Kenny et al., 1994). In contrast, neuroticism and openness to experience are typically judged with lower levels of accuracy and they have fewer valid behavioral cues that are typically available in unstructured experimental situations (Beer & Watson, 2008; Borkenau, Brecke, Möttig, & Paelecke, 2009; Borkenau & Liebler, 1992; John & Robins, 1993; Naumann, Vazire, Rentfrow, & Gosling, 2009). However, neuroticism was judged with significant levels of accuracy when judgments were based on observations of situations that were more relevant to neuroticism

(i.e., socially stressful; Hirschmuller, Egloff, Schmukle, Nestler, & Back, 2015). This last finding highlights the importance of the situation for availability and that whether relevant cues are available for a given trait is likely to depend on the situation.

Which traits are good traits can also depend on who is making the judgment: the self or someone else. Several studies have found evidence of differences in accuracy of behavioral predictions depending on who made the judgments (Kolar, Funder, & Colvin, 1996; Spain, Eaton, & Funder, 2000; Vazire & Mehl, 2008). Vazire (2010) proposed the Self-Other Knowledge Asymmetry (SOKA) model to explain these differences between judgments of the self and judgments of others and to predict which traits and behaviors are likely to exhibit such differences. Two factors were identified as important for differences in judgments of the self vs. others: *perspective*, in terms of access to overt and internal aspects of a person; and *motivation*, in terms of how relevant a judgment is to the ego or the positivity of one's view of the self. Vazire (2010) connects these factors to RAM by specifying that perspective is likely to be relevant to the availability and detection stages, and motivation is likely to be relevant to the detection and utilization stages. Based on the perspective factor, the self typically has more information than others about internal aspects of personality, such as thoughts and feelings and physiological states, and interprets these internal aspects as more salient than external, behavioral aspects of personality. For this reason, the model predicts that self-judgments will be more accurate than other judgments for traits that are manifest in internal thoughts and feelings, such as neuroticism and happiness. Based on the motivational factor, the tendency to see the self and others positively disrupts self-judgments more than other-judgments due to the inability of self-ratings to be based on reality, and therefore self-ratings are predicted to be less accurate for more evaluative traits than are other-ratings. Vazire (2010) found support for the predictions by using self-ratings and other-ratings (friends and strangers) to predict behavior in a laboratory setting.

A fourth moderator of RAM is *good information*, which can be broken down into the aspects of quantity and quality. *Information quantity* is concerned with how many relevant cues are available to judges. A robust finding that relates to information quantity is that judges who have known targets for longer, or who have observed longer recorded interactions, tend to make more accurate judgments than judges who have less exposure to targets (Blackman & Funder, 1998; Connelly & Ones, 2010; Letzring et al., 2006; Paunonen, 1989). This has been labeled the *acquaintanceship effect* (Funder & Colvin, 1988; Watson et al., 2000). Research that has examined how information quantity relates to both distinctive and normative accuracy has shown differences in the relationship

with information quantity. For example, more information is positively related to distinctive accuracy but not to normative accuracy (Biesanz & Human, 2010; Biesanz et al., 2007).

An important aspect of information quantity is that accurate judgments are possible even when there has been little exposure to the target, which is referred to as *zero acquaintance* or *thin slice* research. Research that examines zero acquaintance and thin slices of information has revealed surprising levels of accuracy for even extremely brief exposures (less than 1 minute or 5 minutes) to recorded behavior of targets (Ambady, Hallahan, & Rosenthal, 1995; Ambady & Rosenthal, 1992; Hirschmuller et al., 2015) or photographs of targets (Borkenau & Liebler, 1992; Naumann et al., 2009). Information quantity interacts with the good trait moderator, as zero acquaintance accuracy is typically higher for good (vs. bad) traits. Traits that tend to be judged with especially high levels of accuracy at zero acquaintance include extraversion, positive affect, self-esteem, and intelligence; and traits that tend to be judged less accurately include neuroticism, openness, dominance, and trustworthiness (Borkenau & Liebler, 1992; Hall et al., 2008; Naumann et al., 2009; Rule et al., 2013). Research that examined the influence of information quantity on distinctive and normative accuracy found that distinctive accuracy was higher for targets who spoke more words and when judges paid better attention to targets (Human et al., 2014), both of which imply having more information on which to base judgments. The second factor, attention to the target, was more predictive of accuracy when targets were more (vs. less) well-adjusted, which supports the importance of relevant information for accuracy.

The other aspect of good information is quality, which is directly related to the relevance stage of RAM in that high quality information is more relevant to the trait being judged, and therefore results in more accurate judgments of that trait. Cline and colleagues' (1972) finding that "what most interviewees say in response to interview questions is far more important, as cues, than what they look like, what the voice sounds like, how they act or move (without the sound) all put together" (p. 390) suggests that some types of information are more useful for making accurate judgments than others. This notion is supported by the finding that judges who listened to targets talk about their thoughts and feelings made more accurate judgments than those who listened to targets talk about their hobbies and activities (Andersen, 1984), suggesting that thoughts and feelings are of higher quality than information about hobbies and activities. In another study, an experimental manipulation based on instructions to participants to get to know each other as well as possible, talk about whatever they would like, or discuss trivia questions, revealed higher accuracy in the first two situations, which were designed to elicit more personality-

relevant information, as compared to the other situation (Letzring et al., 2006). There is also evidence that some types of information are especially useful for accurately judging certain traits, such that information about values is especially useful for judgments of neuroticism and information about how a target is unique is especially useful for judgments of conscientiousness (Beer & Brooks, 2011).

The impact of information quality has been shown to differ for distinctive and normative accuracy. Interaction partners who were instructed to talk about their thoughts and feelings or typical behaviors achieved higher levels of distinctive accuracy than interaction partners who were instructed to engage in a series of behaviors, while normative accuracy did not differ across these conditions (Letzring & Human, 2014). Based on codings of the information available in these interactions, distinctive accuracy was higher for openness and ego-control when more information about thoughts and feelings was available, and was higher for extraversion, neuroticism, openness, and ego-control when more information about behaviors was available. Perhaps even more striking, participants who talked more about their own personality, which would certainly be high quality information, were judged with higher levels of distinctive accuracy for extraversion, agreeableness, conscientiousness, neuroticism, and ego-control. Normative accuracy also tended to be higher across traits when more information about thoughts and feelings was available, higher for extraversion when more information about behaviors was available, but lower for agreeableness and conscientiousness when more information about behaviors was available. The negative relationships for information quantity and normative accuracy may be due to judges needing to rely less on their concept of the average person or on self-perceptions when making judgments of their partners. This research also demonstrates the potential overlap between the aspects of information, in that both quantity and type of information can influence accuracy.

As discussed in the section on the good trait, significant levels of accuracy for traits that are typically harder to judge can occur when the situation is relevant to the trait (Hirschmuller et al., 2015). This finding convincingly illustrates the importance of relevant information to accuracy and suggests that accuracy can be increased by attending to the situations in which targets are observed and the information that is likely to be made available in those situations.

The State and Trait Accuracy Model (STAM)

The State and Trait Accuracy Model (STAM) has been proposed as a way to combine research and theorizing about how states – and in particular, affective states – and personality traits are accurately judged (Hall,

Gunnery, Letzring, Carney, & Colvin, in press). STAM builds on RAM by incorporating the stages of RAM in two phases of judgments. First, the stages of RAM are completed to make judgments about affective states based on available cues, and then the judgments about affective states are used, along with other cues, to make judgments of traits. Based on this two-step process, STAM predicts that more accurate judgments of affect will result in more accurate judgments of traits. Preliminary support for the model was found in the positive relations between accuracy of judging distressed affect and neuroticism, and between accuracy of judging positive affect and extraversion(Hall et al., in press). A series of studies is in progress to specifically test this model by first examining the descriptive relationship between affect accuracy and trait accuracy, and then using experimental manipulations to test the causal relationship between accuracy for the two domains.[2]

CUES USED IN JUDGMENTS

Many types of cues can be used to make personality judgments. Lens model analyses code the presence of several cues and examine how those cues relate to the characteristics of the targets and to the ratings made by judges (see Brunswik's Lens Model and the Dual Lens Model section in this chapter). The majority of research on interpersonal accuracy for traits involves many kinds of cues, as judgments are typically made following interactions between judges and targets or observations of recorded behavior of the targets. In this case, judgments will be based on both verbal and nonverbal cues. Verbal cues consist of what the person actually said, which can be coded to determine the amount of information that is available to judges that is relevant to certain topics (see Letzring & Human, 2014). Verbal cues can be isolated from nonverbal cues by only playing the audio portion of the stimulus materials. It is also possible to make only nonverbal cues available by playing recorded behavior without the sound (Borkenau & Liebler, 1992, 1995) or by using only still photographs (Borkenau et al., 2009; Borkenau & Liebler, 1992; Naumann et al., 2009). Nonverbal cues include physical appearance, type of dress, facial features and expressions, tone of voice, speed of speaking, rate and type of movement, and many others (Borkenau & Liebler, 1992, 1995; Koppensteiner, 2013; Naumann et al., 2009). Judges can make relatively accurate judgments of some traits when exposed to only static nonverbal cues (as in a photograph; Borkenau et al., 2009; Naumann et al., 2009), although accuracy of strangers tends to increase with richer stimuli that also include movement and verbal information (Borkenau & Liebler, 1992). There is evidence that more information in the form of appearance and voice can increase accuracy for ratings of extraversion and neuroticism, but more information can decrease accuracy for conscientiousness and openness (Wall, Taylor, Dixon, Conchie, & Ellis, 2013).

A different type of cue on which personality judgments can be based are people's physical environments, such as bedrooms and offices. Cues in such environments include identity claims (e.g., posters, collections of memorabilia) and behavioral residue (e.g., drawings, concert tickets, hiking equipment). Gosling and colleagues (Gosling, Ko, Mannarelli, & Morris, 2002) found that ratings based on offices were accurate for judgments of extraversion, conscientiousness, and openness; and ratings based on bedrooms were accurate for all Big Five traits. For both offices and bedrooms, openness was the most accurately judged trait. The relationships between cue validity and cue utilization where highest for conscientiousness and openness, meaning that for these traits people were more likely to base judgments on valid cues and less likely to base judgments on invalid cues. For offices, the valid cues with the strongest relations with conscientiousness were clean, neat, organized, and (un)cluttered; and the valid cues with the strongest relations to openness were distinctiveness of the spaces and unconventional.

Cues to personality also exist in social media, including personal websites, Facebook profiles, and postings on Twitter. Judgments of strangers based on Facebook profiles (Hall, Pennington, & Lueders, 2014) and other personal websites (Back et al., 2010; Kluemper, Rosen, & Mossholder, 2012; Marcus, Machilek, & Schutz, 2006; Vazire & Gosling, 2004) have achieved accuracy for all Big Five traits; although accuracy is consistently found across studies and ways of computing accuracy only for extraversion and conscientiousness. Furthermore, in one study a relationship was found between accuracy and the distinctiveness of the website (Marcus et al., 2006). Based on lens model analyses, valid cues were utilized to make some judgments (e.g., total number of friends and positive affect in status updates for extraversion, friendliness of profile pictures for agreeableness, support by friends of status updates for conscientiousness, sharing more media for openness; Hall et al., 2014). Judgments of Facebook profiles made by trained evaluators, who were educated in management, information systems, or business administration, agreed with self-ratings for all Big Five traits, and ratings of emotional stability and agreeableness based on the Facebook profiles correlated with job performance ratings from actual supervisors of the targets (Kluemper et al., 2012). Ratings based on 1-month of Twitter posts in an international sample of targets achieved accuracy for agreeableness and neuroticism, but only for aggregate reports of judges and not for single reports (Qiu, Lin, Ramsay, & Yang, 2012).

Cues to personality can also be based on only text, such as that found in essays and creative writing. Borkenau and colleagues (Borkenau, Mosch, Tandler, & Wolf, 2014) had targets write essays on five topics (hobbies, friends, family, academic studies, and plans for the future) and judges made ratings of personality after

reading one or all of the essays. Significant levels of distinctive and normative accuracy were found across all traits, and the highest levels of overall accuracy were found for ratings of openness and for judges who read all essays vs. only one essay. Several cues were coded for each essay (e.g., length, detail, positive emotions), and cue validity and cue utilization were related across all cues for all traits except extraversion. Valid and utilized cues included the following: negative emotion and self-criticism for neuroticism, use of exaggerations and facts reported (negative) for extraversion, elaborate language and influences others for openness, extent of inconsistencies (negative) and negative emotion (negative) for agreeableness, and elaborate language and irony and sarcasm (negative) for conscientiousness. In a similar study, Küfner and colleagues (Küfner, Back, Nestler, & Egloff, 2010) had judges assess targets based on creative writing samples. Significant levels of accuracy were found for openness, agreeableness, and general knowledge. Valid cues that were utilized included creative expression and positive emotions for openness, positive emotions and social orientation for agreeableness, and sophisticated writing and creative expression for knowledge.

Some research has focused on the characteristics of the judges that are related to the ability to accurately judge personality based on text only. Hall and colleagues (Hall, Goh, Schmid Mast, & Hagedorn, 2015) examined the relations between personality traits of judges and accuracy based on reading text excerpts. Relations were found with several judge characteristics, including gender (females were more accurate), agreeableness, conscientiousness, empathic concern, interest in people's personalities, and amount of reading for pleasure.

A meta-analysis of studies that reported accuracy scores for ratings of the Big Five traits that were based on written material from the targets or online social network profiles found significant mean accuracy scores for all traits except neuroticism, and accuracy did not differ for written material vs. online social network profiles (Tskhay & Rule, 2014). Importantly, high levels of heterogeneity were found across the effect sizes, which suggests that moderators of accuracy are likely (although only one moderator was examined due to the small number of studies in the meta-analysis). However, this heterogeneity could also mean that the current findings are unreliable, and therefore more work needs to be done to uncover the true relationships when judgments are based on written material. Overall, there is much research evidence that judges can use many different kinds of cues in the process of judging others, and some cues are more useful for judging some traits.

THE ROLE OF STEREOTYPES AND PROJECTION

Sometimes, there is little information to use when making judgments of others. Judges tend to default to two options when this is the case: they base judgments on stereotypes or their view of themselves. Stereotypes are "beliefs about the attributes of groups and their individual members" (Jussim, Harber, Crawford, Cain, & Cohen, 2005, p. 90). Stereotypes are often thought to be negative, and a good deal of research has focused on the use of inaccurate and negative stereotypes. However, stereotypes can provide valid information about others (Cline et al., 1972; Jussim, Crawford, Harber, & Cohen, 2009; Jussim et al., 2005; Lee, McCauley, & Jussim, 2013). In zero-acquaintance situations, judgments can be based on stereotypes of groups to which the targets belong that can be determined based on the target's physical appearance, such as gender, ethnicity, or approximate age. When there is correspondence between a group stereotype and what a group is actually like, then judgments based on stereotypes will result in at least partially accurate judgments. This fits with Funder's RAM (see the Realistic Accuracy Model section) in that group membership based on physical appearance is available and can be detected by the judge. If there is relevant personality information associated with group membership, then using that information correctly will increase accuracy of judgments. However, if the stereotype is wrong, meaning that information based on the stereotype is not relevant to personality, then using the stereotype will decrease accuracy. There is evidence that stereotypes, both those that are shared by a culture and specific to an individual, have surprisingly large amounts of valid information and are accurate 70-90% of the time (Lee et al., 2013), which means that using stereotypes can result in accurate judgments (Jussim et al., 2005). Even with this high level of validity, "stereotypes have only a very small influence on person perception" (Lee et al., 2013, p. 477), and are more likely to be used when judges have little information about the targets.

Another way to make judgments when little information is available about targets is to base judgments on how one sees him/herself, which is referred to as *assumed similarity*, a *self-based heuristic*, or *projection* (Cronbach, 1955; Ready et al., 2000; Watson et al., 2000). This method is likely to be used when judgments are about internal and therefore less observable items or traits, or for traits for which judges have little individuating information (Beer & Watson, 2008; Paunonen & Hong, 2013; Watson et al., 2000). It is also possible for projection to be used in ratings of well-acquainted individuals, and when this occurs accuracy is facilitated when the judge and target are actually similar to each other on the characteristic being judged (Paunonen & Hong, 2013; Ready et al., 2000; Watson et al., 2000). Projection sometimes reflects a self-bias and sometimes it reflects accurate knowledge. The

bias occurs when the judge and target are not similar on the attribute being judged, but the judge uses self-perceptions to make judgments.

META-ACCURACY

Meta-accuracy is the accuracy of "our beliefs about how others perceive us" (Carlson, 2016a, p. 717). This topic fits within the umbrella of person perception accuracy, but differs from other work in this area in that the focus is on how accurately people know how others perceive them, rather than on how accurately people perceive others. For this reason, only a brief overview of this area will be presented here. Among strangers and acquaintances, people can achieve significant levels of meta-accuracy, meaning that they are accurate about how others' perceive their own personality traits (Carlson, Furr, & Vazire, 2010; Carlson, Vazire, & Furr, 2011; Kenny & DePaulo, 1993; Shechtman & Kenny, 1994), even when controlling for the normativeness of judgments (Carlson & Furr, 2009). Early research on this topic found that the magnitude of meta-accuracy tended to be higher for judgments of perceptions made by others in general vs. by a specific other (Kenny & DePaulo, 1993; Levesque, 1997; Shechtman & Kenny, 1994). However, some research has shown that people can have differentiated meta-accuracy for relationships from different contexts (e.g., parents, college friends; Carlson & Furr, 2009). Meta-accuracy also seems to be an individual difference, based on the finding of significant variance of meta-accuracy across individuals, and people tend to know how accurate their meta-perceptions are (Carlson & Furr, 2009, 2013). Some characteristics of people are related to meta-accuracy. For example, people who are socially anxious tend to be accurate in believing that others see them negatively and people high in need for approval tend to be accurate in believing that others like them (Kenny & DePaulo, 1993), and people who are psychologically well-adjusted tend to project more of their own self-perceptions onto their meta-perceptions, compared to less well-adjusted people (Mosch & Borkenau, 2016). Also, meta-accuracy tends to be higher in dyads with higher liking for each other (Ohtsubo, Takezawa, & Fukuno, 2009), although the relations between meta-accuracy and relationship quality are complex and differ for liking of the judge vs. the meta-perceiver (Carlson, 2016b).

CORRELATES OF ACCURACY

Several characteristics of judges that correlate with how accurately they tend to judge others were described in the section on the Realistic Accuracy Model in which the good judge moderator was described. Even though the research upon which these findings are based is largely correlational in nature, it is generally reasonable to assume – although not conclusively provable – that the characteristics of the judge are causal mechanisms of accuracy. For

example, it is assumed that agreeable judges cause targets to feel more comfortable and therefore emit more relevant cues to personality, which increases accuracy (Letzring, 2008). A study in which an attribute of a judge that is assumed to affect accuracy is manipulated, and the effect on accuracy assessed, would provide evidence regarding the causal direction of the relationship, and this is an important avenue for future research. However, it is difficult to manipulate characteristics and traits (although not impossible if one focuses on behavior over a limited time; see McNiel, Lowman, & Fleeson, 2010), and this may be why this research has not yet been conducted and reported. On the other hand, it is possible to think of likely outcomes of interpersonal accuracy and to test whether these outcomes are related to accuracy for judging traits. Using this approach, many interesting relations have been uncovered.

One important outcome that has been studied is relationship satisfaction or quality. Self-other agreement among romantic couples has been found to predict relationship quality in several studies, and even has incremental validity beyond self- and partner-ratings of personality (Burke & Stets, 1999; Decuyper, De Bolle, & De Fruyt, 2012; Letzring & Noftle, 2010; Swann, De La Rone, & Hixon, 1994). Among newlyweds, seeing one's partner accurately predicts both one's own satisfaction with the relationship as well as the partner's satisfaction, and this prediction is above and beyond the effects of a similarity bias (believing the partner is more similar to the self than is true) and a positivity bias (seeing the partner in an overly positive way; Luo & Snider, 2009). Some evidence indicates that the relationship between accuracy and relationship satisfaction is stronger among married couples than among unmarried couples (Letzring & Noftle, 2010). Further research has found that wives with higher accuracy provided more support for husbands during a laboratory interaction, felt more control within the marriage after six months, and were less likely to get divorced in the first two years or four years of marriage (even when controlling for initial satisfaction and the positivity of perceptions). For some reason, the same effects were not found for husbands (Neff & Karney, 2005). However, at least one study has shown that self-other agreement among married couples only predicted relationship quality when self- and partner-ratings of personality were not in the model (Furler, Gomez, & Grob, 2014) and another study did not find consistent evidence of a link between accuracy and relationship well-being – in terms of a composite of adjustment, intimacy, and trust – among newlyweds (Pollmann & Finkenauer, 2009).

Research has also examined variables that predict relationship outcomes for new acquaintances. Judges who achieved higher levels of distinctive accuracy at the beginning of a semester interacted more with their partners

during the semester, and had higher liking for their partners and more interest in future interactions at the end of the semester (Human, Sandstrom, Biesanz, & Dunn, 2012). Judges who achieved higher levels of normative accuracy, which indicates positive perceptions of others, at the beginning of the semester, liked their partners better and were more interested in talking again in the future. Normative accuracy also predicted more interaction during the semester, greater liking, greater interest in future interactions, and establishment of a social tie at the end of the semester.

In addition to relationship outcomes, judgments based on observations of recorded interactions have provided evidence of relationships with other interpersonal variables. For example, normative accuracy for judgments of some traits was positively related with self-reports of interpersonal control and interpersonal support (Letzring, 2015). Somewhat surprisingly, distinctive accuracy was not related to interpersonal outcomes. This lack of relationship may be due to judgments being based on observations rather than interaction, judges being unacquainted with the targets, or to less variability in distinctive accuracy across judges as compared to the variability in normative accuracy. It could also reflect the fact, mentioned earlier, that accurate judgments of (statistically) normal people do not contribute to distinctive accuracy scores, yet such judgments are important because, by definition, most of the people one encounters are "normal." Based on other findings described in this section, it is likely that distinctive accuracy is more related to interpersonal outcomes when judges are better acquainted with the targets.

IMPROVING ACCURACY

Based on the many favorable correlates and positive outcomes of interpersonal accuracy, the logical next questions are whether accuracy can be improved, and if so what are the most effective types of training? The most common types of training across domains of person perception are instruction, practice, and feedback (Blanch-Hargian, Andrzejewski, & Hill, 2012). A meta-analysis of training studies for judgments of internal states or traits based on verbal and/or nonverbal cues found a positive and statistically significant, although small, mean effect size of training that was heterogeneous across studies (Blanch-Hargian et al., 2012). The effect was smallest for the domain to which judgments of traits belonged (which also included judgments of status and rapport). Training was most effective when it included practice or feedback, was with individuals or small groups rather than large groups, and when there was a trainer present vs. computer-mediated training. Effectiveness was unrelated to the length of the intervention.

Across studies for which personality traits were the domain of judgment, evidence for the effectiveness of training is mixed. In three studies, Hall and colleagues (Hall et al., 2009) found no effects for several strategies of increasing motivation on judgments of dominance and extraversion. Deborah Powell and colleagues have designed and tested a few interventions aimed at increasing accuracy based on observations of mock job interviews. In two studies, judges' ratings were compared to self-ratings and an average of ratings by ten "experts." Powell and Goffin (2009) tested the effectiveness of a 3-part training in which judges listened to a lecture about personality traits, completed a written exercise to practice using personality-relevant cues to make judgments, and discussed their ratings and received feedback. For distinctive accuracy, which was calculated as a profile correlation across six items, training was not effective for increasing accuracy. When difference scores were used to calculate accuracy, ratings of two out of three characteristics that were the focus of the training - assertiveness and self-discipline - were more accurate for judges in the training group as compared to the control group. There was not a difference between the training and control group for ratings of vulnerability to stress. Furthermore, for comparisons to expert-ratings only, ratings of two out of three characteristics that were not the focus of training but related to the same Big Five traits - cheerfulness and order - were also more accurate for judges in the training group as compared to the control group, and there was not a difference between the training and control group for the third characteristic – self-consciousness. Powell and Bourdage (2016) designed interventions that focused on the last two stages of RAM: detection and utilization. Accuracy based on profile correlations of judgments of the Big Five traits was higher for utilization training and for utilization and detection training together than for detection training alone and for no training. The effects of training were less consistent for item-level accuracy for specific traits. For extraversion and neuroticism, some types of training were more effective than no training; but there were no effects of training for agreeableness (accuracy was low for all conditions), or for openness and conscientiousness (accuracy was high for all conditions). The overall conclusion was that training judges to focus on utilization of cues was the most effective type of training, although this might have also taught judges to detect certain cues.

SUMMARY AND FUTURE DIRECTIONS

An excellent base of knowledge has been established for research on interpersonal accuracy for judgments of personality traits, and empirical evidence exists to support theories of the process and moderators of judgment. There are many opportunities to continue to build on existing knowledge and to forge new connections with related areas of interpersonal accuracy research.

One future direction is to examine a larger variety of personality traits to test the generalizability and specificity of current knowledge. Much of the more current work in this area has focused on the Big Five personality traits, although some researchers have examined other traits such as trustworthiness(Rule et al., 2013) (Rule et al., 2013) and perceptions of and attitudes toward risk (Mishra & Sritharan, 2012; Vineyard, 2016), and concepts similar to traits such as personal values (McDonald & Letzring, in press). Research about domains beyond the Big Five would broaden our understanding of the cues that are important for judging certain traits and is likely to increase the applicability of this area of research. In addition to examining other traits, it would also be useful to examine how accurately judges could rate the variability of targets in their state-level expression of traits. Fleeson and colleagues (Fleeson, 2001; Fleeson & Jayawickreme, 2015; Fleeson & Law, 2015) have presented convincing evidence that peoples' levels of trait-relevant states vary over time and also have a meaningful average. We are not aware of any work that has examined the accuracy with which people can judge variability of trait-relevant states, but this could be an important avenue for new research as such an understanding would reflect knowledge of others at a much deeper level than just knowing their average trait levels.

A second future direction would further examine the connection between accuracy of judging states and accuracy of judging traits. Letzring and Hall (Hall et al., in press) have proposed the State and Trait Accuracy Model (STAM) as a way to combine these areas of accuracy research (see earlier section on the STAM in this chapter). Several studies have examined the correlations between judgments of different domains, but relationships between domains are inconsistent (Ambady et al., 1995; Boone, Schlegel, & Hall, 2016; Hall et al., in press; Realo et al., 2003). A model that proposes that accurate state judgments are a necessary first step – whether judges are aware of it or not – in the process of making accurate trait judgments will lead to novel predictions and empirical tests of those predictions that are likely to increase our understanding of the process of accurate trait judgments. Training programs and interventions based on this increased understanding are likely to be successful at increasing accuracy of both types of judgments. Further research on training or improving accuracy, whether based on results of research that examine relations between accuracy of judging states and traits or not, would be useful for clarifying the current state of research on this topic that has shown mixed results (see section on Improving Accuracy). Training programs could be aimed at a variety of populations, including medical professionals, teachers, clinical psychologists, and lay persons. In fact, training programs might need to be specifically tailored for certain populations to maximize effectiveness.

A third important direction is to create a standardized, reliable, and well-validated assessment or test of the ability to accurately judge personality traits of others. Such tests exist for accuracy for judging other domains, especially based on nonverbal cues (Buck, 1976; Costanzo & Archer, 1989; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), but a standardized test does not exist for judging personality traits. At least a few researchers have noted the usefulness that such a test would have (Boone et al., 2016; Murphy, 2016), and we whole-heartedly agree with them. A standardized test would be especially useful for increasing the pace at which research could take place and would increase the comparability and replicability of results.

A fourth important direction is to expand our knowledge of accuracy in other age groups and the developmental trajectory of the ability to make accurate judgments of the traits of others. The majority of research in this area uses college students and young adults as both judges and targets. Some research has examined the effects of age on emotion recognition. Older adults can identify emotions in prototypical facial expressions as well as young adults for some emotions (e.g., happy, surprise, neutral), although older adults were less accurate than young adults for judgments of fear and sadness (Keightley, Winocur, Burianova, Hongwanishkul, & Grady, 2006), or for fear, sadness, and surprise (Orgeta, 2010). Whether accuracy differs across age groups has also been shown to depend on the difficulty of the recognition task (Orgeta, 2010). However, we are not aware of any research, at present, on relationships between age and accuracy of judgments of personality traits.

In summary, much has been learned about interpersonal accuracy for personality traits and ample evidence indicates that people can and do make accurate judgments, at least some of the time. The level of accuracy that is attained depends on several factors and on the type of accuracy that is assessed. There is still debate about whether interpersonal accuracy is a general skill that cuts across domains and types of judgments, or whether it is a skill that is domain-specific and dependent on how accuracy is calculated. Whether accuracy is general or specific, we know that it is related to many beneficial characteristics of judges and targets, and also seems to lead to beneficial outcomes. The more we can learn about the processes that must take place for accurate judgments to be made, the more likely it is that we will create effective interventions to increase accuracy and, in turn, increase beneficial interpersonal outcomes.

References

Adams, H. F. (1927). The good judge of personality. *Journal of Abnormal and Social Psychology, 22*, 172-181.

Allport, G. W. (1937). Personality: A psychological interpretation. New York, NY: Holt, Reinhart, & Winston.

Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology, 69*, 518-529.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256-274.

Andersen, S. M. (1984). Self-knowledge and social inference: II. The diagnosticity of cognitive/affective and behavioral data. *Journal of Personality and Social Psychology, 46*, 294-307.

Back, M. D., & Nestler, S. (2016). Judging personality. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately*. Cambridge, United Kingdom: Cambridge University Press.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science, 21*, 372-374. doi:10.1177/0956797609360756

Banziger, T. (2016). Accuracy of judging emotions. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 23-51). Cambridge, United Kingdom: Cambridge University Press.

Barenbaum, N. B., & Winter, D. G. (2008). History of modern personality theory and research. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research, 3rd ed.* (pp. 3-26). New York, NY: The Guilford Press.

Beer, A. (2013). Group personality judgments at zero acquaintance: Communication among judges versus aggregation of independent evaluations. *Journal of Research in Personality, 47*, 385-389. doi:10.1016/j.jrp.2013.03.008

Beer, A., & Brooks, C. (2011). Information quality in personality judgment: The value of personal disclosure. *Journal of Research in Personality, 45*, 175-185. doi:10.1016/j.jrp.2011.01.001

Beer, A., & Watson, D. (2008). Personality judgment at zero acquaintance: Agreement, assumed similarity, and implicit simplicity. *Journal of Personality Assessment, 90*, 250-260. doi: 10.1080/00223890701884970

Beer, A., & Watson, D. (2010). The effects of information and exposure on self-other agreement. *Journal of Research in Personality, 44*, 38-45. doi:10.1016/j.jrp.2009.10.002

Biesanz, J. C., & Human, L. J. (2010). The cost of forming more accurate impressions: Accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal. *Psychological Science, 21*, 589-594. doi:10.1177/0956797610364121

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology, 92*, 119-135. doi:10.1037/0022-3514.92.1.119

Blackman, M. C., & Funder, D. C. (1998). The effect of information on consensus and accuracy in personality judgment. *Journal of Experimental Social Psychology, 34*, 164-181.

Blanch-Hargian, D., Andrzejewski, S. A., & Hill, K. M. (2012). The effectiveness of training to improve person perception accuracy: A meta-analysis. *Basic and Applied Social Psychology, 34*, 483-498. doi:10.1080/01973533.2012.728122

Boone, R. T., Schlegel, K., & Hall, J. A. (2016). Individual differences in interpersonal accuracy: Using meta-analysis to assess whether judging other people is one skill or many. *Manuscript submitted for publication*.

Borkenau, P., Brecke, S., Möttig, C., & Paelecke, M. (2009). Extraversion is accurately perceived after a 50-ms exposure to a face. *Journal of Research in Personality, 43*, 703-706. doi:10.1016/j.jrp.2009.03.007

Borkenau, P., & Liebler, A. (1992). Trait inferences: Soruces of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62*, 645-657.

Borkenau, P., & Liebler, A. (1995). Observable attributes as manifestations and cues of personality and intelligence. *Journal of Personality, 63*, 1-25.

Borkenau, P., Mosch, A., Tandler, N., & Wolf, A. (2014). Accuracy of judgments of personality based on textual information on major life domains. *Journal of Personality, 84*, 214-224. doi:10.1111/jopy.12153

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193-217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.

Buck, R. (1976). A test of nonverbal receiving ability: Preliminary studies. *Human Communication Research, 2*, 162-171. doi:10.1111/j.1468-2958.1976.tb00708.x

Burgoon, J. K., & Dunbar, N. E. (2016). Accuracy of distinguishing truth from lie. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately.* (pp. 71-97). Cambridge, United Kingdom: Cambridge University Press.

Burke, P. J., & Stets, J. E. (1999). Trust and commitment through self-verification. *Social Psychology Quarterly, 62*, 347-366.

Carlson, E. N. (2016a). Do psychologically adjusted individuals know what other people really think about them? The link between psychological adjustment and meta-accuracy. *Social Psychological and Personality Science, 7*, 717-725. doi:10.1177/1948550616646424

Carlson, E. N. (2016b). Meta-accuracy and relationship quality: Weighing the costs and benefits of knowing what people really think about you. *Journal of Personality and Social Psychology, 111*, 250-264. doi:10.1037/pspp0000107

Carlson, E. N., & Furr, R. M. (2009). Evidence of differential meta-accuracy. *Psychological Science, 20*, 1033-1039.

Carlson, E. N., & Furr, R. M. (2013). Resolution of meta-accuracy: Should people trust their beliefs about how others see them? *Social Psychological and Personality Science, 4*, 419-426. doi:10.1177/1948550612461653

Carlson, E. N., Furr, R. M., & Vazire, S. (2010). Do we know the first impressions we make? Evidence for idiographic meta-accuracy and calibration of first impressions. *Social Psychological and Personality Science, 1*, 94-98. doi:10.1177/1948550609356028

Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-insight: Do people really know how others see them? *Journal of Personality and Social Psychology, 101*, 831-846. doi:10.1037/a0024297

Cline, V. B., Atzet, J., & Holmes, E. (1972). Assessing the validity of verbal and nonverbal cues in accurately judging others. *Comparative Group Studies, 3*, 383-394.

Colman, D. E., Letzring, T. D., & Biesanz, J. C. (2016). Seeing and feeling your way to accurate personality judgments: The moderating role of perceiver empathic tendencies. *Manuscript submitted for publication*.

Colvin, C. R. (1993). “Judgable” people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology, 64*, 861-873.

Connelly, B. S., & Ones, D. S. (2010). An other perspetive on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*, 1092-1122. doi:10.1037/a0021212

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior, 13*, 225-245. doi: 10.1007/BF00990295

Cronbach, L. J. (1955). Processes affecting scores on “understanding of others” and “assumed similarity”. *Psychological Bulletin, 52*, 177-193.

Decuyper, M., De Bolle, M., & De Fruyt, F. (2012). Personality similarity, perceptual accuracy, and relationship satisfaction in dating and married couples. *Personal Relationships, 19*, 128-145. doi:10.1111/j.1475-6811.2010.01344.x

Estes, S. G. (1938). Judging personality from expressive behavior. *The Journal of Abnormal and Social Psychology, 33*, 217-236. doi:10.1037/h0058565

Finkel, E. J., & Eastwick, P. W. (2008). Speed-dating. *Current Directions in Psychological Science, 17*, 193-197. doi:10.1111/j.1467-8721.2008.00573.x

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*, 1011-1027. doi:10.1037//0022-3514.80.6.1011

Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality, 56*, 82-92. doi:10.1016/j.jrp.2014.10.009

Fleeson, W., & Law, M. K. (2015). Trait enactments as density distributions: The role of actors, situations, and observers in explaining stability and variability. *Journal of Personality and Social Psychology, 109*, 1090-1104. doi:10.1037/a0039517

Funder, D. C. (1980). On seeing ourselves as other see us: Self-other agreement and discrepancy in personality ratings. *Journal of Personality, 48*, 473-493.

Funder, D. C. (1983). The "consistency" controversy and the accuracy of personality judgments. *Journal of Personality, 51*, 346-359.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101*, 75-90.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652-670.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science, 21*, 177-182. doi:10.1177/0963721412445309

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*, 149-158.

Funder, D. C., & Colvin, C. R. (1997). Congruence of self and others' judgments of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology.* (pp. 617-647). Orlando, FL: Academic Press.

Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52*, 409-418.

Furler, K., Gomez, V., & Grob, A. (2014). Personality perceptions and relationship satisfaction in couples. *Journal of Research in Personality, 50*, 33-41. doi:10.1016/j.jrp.2014.02.003

Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality, 76*, 1267-1316. doi:10.1111/j.1467-6494.2008.00521.x

Gage, N. L., & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review, 62*, 411-422.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*, 451-482.

Gilbert, D. T. (1988). Ordinary Personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology, Vols. 1 and 2.* (4 ed., pp. 89-150). New York, NY: McGraw-Hill.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*, 148-168. doi:10.1037/a0034726

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology, 82*, 379-398. doi:10.1037//0022-3514.82.3.379

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Schmid Mast, M., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality, 42*, 1476-1489. doi:10.1016/j.jrp.2008.06.013

Hall, J. A., Blanch, D. C., Horgan, T. G., Murphy, N. A., Rosip, J. C., & Schmid Mast, M. (2009). Motivation and interpersonal sensitivity: Does it matter how hard you try? *Motivation and Emotion, 33*, 291-302. doi:10.1007/s11031-009-9128-2

Hall, J. A., & Colvin, C. R. (2011). On making a thriving field even better: Acknowledging the past and looking to the future. *Psychological Inquiry, 22*, 193-199. doi: 10.1080/1047840X.2011.573765

Hall, J. A., Goh, J. X., Schmid Mast, M., & Hagedorn, C. (2015). Individual differences in accurately judging personality from text. *Journal of Personality, 84*, 433-445. doi: 10.1111/jopy.12170

Hall, J. A., Gunnery, S. D., Letzring, T. D., Carney, D. R., & Colvin, C. R. (in press). Accuracy of judging affect and accuracy of judging personality: How and when are they related? *Journal of Personality*. doi: 10.1111/jopy.12262

Hall, J. A., Pennington, N., & Lueders, A. (2014). Impression management and formation on Facebook: A lens model approach. *New Media and Society, 16*, 958-982. doi: 10.1177/1461444813495166

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62*, 255-262.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71*, 438-456.

Hartwig, M., & Bond, C. F., Jr. (2014). Lie detection from multiple cues: A meta-analysis. *Applied Cognitive Psychology, 28*, 661-676. doi:10.1002/acp.3052

Harvey, J. H., Town, J. P., & Yarkin, K. L. (1981). How fundamental is "The Fundamental Attribution Error"? *Journal of Personality and Social Psychology, 40*, 346-349.

Hastie, R., & Rasinski, K. A. (1988). The concept of accuracy in social judgment. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social psychology of knowledge.* (pp. 193-208). New York: NY: Cambridge University Press.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York, NY: Wiley.

Hirschmuller, S., Egloff, B., Nestler, S., & Back, M. D. (2013). The dual lens model: A comprehensive framework for understanding self-other agreement of personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 104*, 335-353. doi:10.1037/a0030383

Hirschmuller, S., Egloff, B., Schmukle, S. C., Nestler, S., & Back, M. D. (2015). Accurate judgments of neuroticism at zero acquaintance: A question of relevance. *Journal of Personality, 83*, 221-228. doi:10.1111/jopy.12097

Hodges, S. D., Lewis, K. L., & Ickes, W. (2015). The matter of other minds: Empathic accuracy and the factors that influence it. In P. Shaver, M. Mikulincer, J. A. Simpson, & J. Dovidio (Eds.), *APA handbook of personality and social psychology, Vol 3. Interpersonal relations.* (pp. 319-348). Washington, DC: American Psychological Association.

Human, L. J., Biesanz, J. C., Finseth, S. M., Pierce, B., & Le, M. (2014). To thine own self be true: Psychological adjustment promotes judgeability via personality-behavior congruence. *Journal of Personality and Social Psychology, 106*, 286-303. doi:10.1037/a0034860

Human, L. J., Sandstrom, G. M., Biesanz, J. C., & Dunn, E. W. (2012). Accurate first impressions leave a lasting impression: The long-term effects of distinctive self-other agreement on relationship development. *Social Psychological and Personality Science, 4*, 395-402. doi:10.1177/1948550612463735

Ickes, W. (2016). Empathic accuracy: Judging thoughts and feelings. In J. A. Hall, M. Schmid Mast, & S. G. West (Eds.), *The social psychology of perceiving others accurately.* (pp. 52-70). Cambridge, United Kingdom: Cambridge University Press.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*, 521-551.

Jussim, L., Crawford, J. T., Harber, K., & Cohen, F. (2009). The unbearable accuracy of stereotypes. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination.* (pp. 199-227). New York, NY: Psychology Press.

Jussim, L., Harber, K. D., Crawford, J. T., Cain, T. R., & Cohen, F. (2005). Social reality makes the social mind: Self-fulfilling prophecy, stereotypes, bias, and accuracy. *Interaction Studies, 6*, 85-102.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.

Kammann, R., Smith, R., Martin, C., & McQueen, M. (1984). Low accuracy in judgments of others' psychological well-being as seen from a phenomenological perspective. *Journal of Personality, 52*, 107-123.

Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin, 134*, 404-426. doi:10.1037/0033-2909.134.3.404

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology.* (pp. 451-477). New York, NY: Cambridge University Press.

Keightley, M. L., Winocur, G., Burianova, H., Hongwanishkul, D., & Grady, C. L. (2006). Age effects on social cognition: Faces tell a different story. *Psychology and Aging, 21*, 558-572. doi: 10.1037/0882-7974.21.3.558

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation, Vol. 15.* (pp. 192-238). Lincoln, NE: University of Nebraska Press.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28*, 107-128.

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review, 98*, 155-163.

Kenny, D. A. (1994). *Interpersonal Perception*. New York, NY: The Guilford Press.

Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review, 8*, 265-280.

Kenny, D. A., Albright, L., Malloy, T. E., & Kashy, D. A. (1994). Consensus in interpersonal perception: Acquaintance and the Big Five. *Psychological Bulletin, 116*, 245-258.

Kenny, D. A., & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin, 114*, 145-161.

Kenny, D. A., & Kashy, D. A. (2014). The design and analysis of data from dyads and groups. In H. T. Reis, C. M. Judd, H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology, 2nd ed.* (pp. 589-607). New York, NY, US: Cambridge University Press.

Kenny, D. A., & la Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology, 48*, 339-348. doi:10.1037/0022-3514.48.2.339

Kluemper, D. H., Rosen, P., & Mossholder, K. W. (2012). Social networking websites, personality ratings, and the organizational context: More than meets the eye? *Journal of Applied Social Psychology, 42*, 1143-1172. doi:10.1111/j.1559-1816.2011.00881.x

Kolar, D. W. (1996). *Individual differences in the ability to accurately judge the personality characteristics of others.* (Unpublished doctoral dissertation), University of California, Riverside, Riverside, CA.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality, 64*, 311-337.

Koppensteiner, M. (2013). Motion cues that make an impression: Predicting perceived personality by minimal motion information. *Journal of Experimental Social Psychology, 49*, 1137-1143. doi: 10.1016/j.jesp.2013.08.002

Küfner, A. C. P., Back, M. D., Nestler, S., & Egloff, B. (2010). Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing. *Journal of Research in Personality, 44*, 427-435. doi:10.1016/j.jrp.2010.05.003

Lee, Y., McCauley, C., & Jussim, L. (2013). Stereotypes as valid categories of knowledge and human perceptions of group differences. *Social and Personality Psychology Compass, 7*, 470-486. doi: 10.1111/spc3.12039

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of Research in Personality, 42*(4), 914-932. doi:10.1016/j.jrp.2007.12.003

Letzring, T. D. (2015). Observer judgmental accuracy of personality: Benefits related to being a good (normative) judge. *Journal of Research in Personality, 54*, 51-60. doi:10.1016/j.jrp.2014.05.001

Letzring, T. D., & Human, L. J. (2014). An examination of information quality as a moderator of accurate personality judgment. *Journal of Personality, 82*, 440-451. doi:10.1111/jopy.12075

Letzring, T. D., & Noftle, E. E. (2010). Predicting relationship quality from self-verification of broad personality traits among romantic couples. *Journal of Research in Personality, 44*, 353-362. doi:10.1016/j.jrp.2010.03.008

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology, 91*, 111-123. doi:10.1037/0022-3514.91.1.111

Levesque, M. J. (1997). Meta-accuracy among acquainted individuals: A social relations analysis of interpersonal perception and metaperception. *Journal of Personality and Social Psychology, 72*, 66-74.

Levesque, M. J., & Kenny, D. A. (1993). Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. *Journal of Personality and Social Psychology, 65*, 1178-1187. doi:10.1037/0022-3514.65.6.1178

Luo, S., & Snider, A. G. (2009). Accuracy and biases in newlyweds' perceptions of each other. *Psychological Science, 20*, 1332-1339.

Marcus, B., Machilek, F., & Schutz, A. (2006). Personality in cyberspace: Personal web sites as media for personality expressions and impressions. *Journal of Personality and Social Psychology, 90*, 1014-1031. doi:10.1037/0022-3514.90.6.1014

McDonald, J. S., & Letzring, T. D. (in press). Accuracy of judging personal values and personality traits. *Journal of Research in Personality*.

Mishra, S., & Sritharan, R. (2012). Personality and behavioral outcomes associted with risk-taking are accurately inferred from faces. *Journal of Research in Personality, 46*, 760-764.

Mosch, A., & Borkenau, P. (2016). Psychologically adjusted persons are less aware of how they are perceived by others. *Personality and Social Psychology Bulletin, 42*, 910-922. doi:10.1177/0146167216647383

Murphy, N. A. (2016). What we know and the future of interpersonal accuracy research. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately.* (pp. 404-424). Cambridge, United Kingdom: Cambridge University Press.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin, 35*, 1661-1671. doi:10.1177/0146167209346309

Neff, L. A., & Karney, B. R. (2005). To know you is to love you: The implications of global adoration and specific accuracy for marital relationships. *Journal of Personality and Social Psychology, 88*, 480-497. doi:10.1037/0022-3514.88.3.480

Nestler, S., & Back, M. D. (2013). Applications and extensions of the lens model to understand interpersonal judgments at zero acquaintance. *Current Directions in Psychological Science, 22*, 374-379. doi:10.1177/0963721413486148

Nestler, S., & Back, M. D. (2015). Using cross-classified structural equation models to examine the accuracy of personality judgments. *Psychometrika*. doi:10.1007/s11336-015-9485-6

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. New York, NY: Prentice Hall.

Ohtsubo, Y., Takezawa, M., & Fukuno, M. (2009). Mutual liking and meta-perception accuracy. *European Journal of Social Psychology, 39*, 707-718. doi:10.1002/ejsp.568

Orgeta, V. (2010). Effects of age and task difficulty on recognition of facial affect. *Journal of Gerontology: Psychological Sciences, 65B*, 323-327. doi:10.1093/geronb/gbq007

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57*, 401-421. doi: 10.1146/annurev.psych.57. 102904.190127

Paunonen, S. V. (1989). Consensus in personality judgments: Moderating effects of target-rater acquaintance and behavior observability. *Journal of Personality and Social Psychology, 56*, 823-833.

Paunonen, S. V., & Hong, R. Y. (2013). The many faces of assumed similarity in perceptions of personality. *Journal of Research in Personality, 47*, 800-815. doi: 10.1016/j.jrp.2013.08.007

Paunonen, S. V., & Kam, C. (2014). The accuracy of roommate ratings of behaviors versus beliefs. *Journal of Research in Personality, 52*, 55-67. doi:10.1016/j.jrp.2014.07.006

Pollmann, M. M. H., & Finkenauer, C. (2009). Investigating the role of two types of understanding in relationship well-being: Understanding is more important than knowledge. *Personality and Social Psychology Bulletin, 35*, 1512-1527. doi: 10.1177/0146167209342754

Powell, D. M., & Bourdage, J. S. (2016). The detection of personality traits in employment interviews: Can "good judges" be trained? *Personality and Individual Differences, 94*, 194-199. doi: 10.1016/j.paid.2016.01.009

Powell, D. M., & Goffin, R. D. (2009). Assessing personality in the employment interview: The impact of training on rater accuracy. *Human Performance, 22*, 450-465. doi: 10.1080/08959280903248450

Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality, 46*, 710-718. doi: 10.1016/j.jrp.2012.08.008

Ready, R. E., Clark, L. A., Watson, D., & Westerhourse, K. (2000). Self- and peer-reported personality: Agreement, trait ratability, and the "self-based heuristic". *Journal of Research in Personality, 34*, 208-224. doi:10.1006/jrpe.1999.2280

Realo, A., Allik, J., Nõlvak, A., Valk, R., Ruus, T., Schmidt, M., & Eilola, T. (2003). Mind-reading ability: Beliefs and performance. *Journal of Research in Personality, 37*, 420-445. doi:10.1016/s0092-6566(03)00021-7

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality. *Perspectives on Psychological Science, 2*, 313-345.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS test*. Baltimore, MD: Johns Hopkins University Press.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology, 10*, 173-220. doi:10.1016/S0065-2601(08)60357-3

Ross, L., Amabile, T. M., & Steinmetz, J. L. (1977). Social roles, social control, and biases in social-perception processes. *Journal of Personality and Social Psychology, 35*, 485-494.

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: Behavioral and neural correlates. *Journal of Personality and Social Psychology, 104*, 409-426.

Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). *Person perception* (2nd ed.). Reading, MA: Addison-Wesley.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*, 609-612. doi:10.1016/j.jrp.2013.05.009

Shechtman, Z., & Kenny, D. A. (1994). Metaperception accuracy: An Israeli study. *Basic and Applied Social Psychology, 15*, 451-465.

Spain, J. S., Eaton, L. G., & Funder, D. C. (2000). Perspectives on personality: The relative accuracy of self versus others for the prediction of emotion and behavior. *Journal of Personality, 68*, 837-867.

Swann, W. B., Jr., De La Rone, C., & Hixon, J. G. (1994). Authenticity and positivity strivings in marriage and courtship. *Journal of Personality and Social Psychology, 66*, 857-869.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52*, 1-23.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review, 93*, 239-257.

Trope, Y. (1989). Levels of inference in dispositional judgment. *Social Cognition, 7*, 296-314.

Tskhay, K. O., & Rule, N. O. (2014). Perceptions of personality in text-based media and OSN: A meta-analysis. *Journal of Research in Personality, 49*, 25-30. doi: 10.1016/j.jrp.2013.12.004

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207-232. doi:10.1016/0010-0285(73)90033-9

Vazire, S. (2010). Who knows what about a person? The Self-Other Knowledge Asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*, 281-300. doi:10.1037/a0017908

Vazire, S., & Gosling, S. D. (2004). E-perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology, 87*, 123-132. doi:10.1037/0022-3514.87.1.123

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*, 1202-1216. doi:10.1037/a0013314

Vernon, P. E. (1933). Some characteristics of the good judge of personality. *Journal of Social Psychology, 4*, 42-57.

Vineyard, J. (2016). *The accuracy of juding risk in others: Investigating information quality and individual differences as moderators of accuracy of risk trait judments.* (Unpublished doctoral dissertation), Idaho State University, Pocatello, Idaho.

Vogt, D. S., & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality and Social Psychology, 71*, 267-295.

Wall, H. J., Taylor, P. J., Dixon, J., Conchie, S. M., & Ellis, D. A. (2013). Rich contexts do not always enrich the accuracy of personality judgments. *Journal of Experimental Social Psychology, 49*, 1190-1195. doi:10.1016/j.jesp.2013.05.010

Warner, R. M., Kenny, D. A., & Stoto, M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology, 37*, 1742-1757. doi:10.1037/0022-3514.37.10.1742

Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology, 78*, 546-558. doi:10.I037//O022-3514.78.3.546

Wood, D., & Furr, R. M. (2016). The correlates of similarity estimates are often midleadingly positive: The nature and scope of the proglem, and some solutions. *Personality and Social Psychology Review, 20*, 79-99. doi:10.1177/1088868315581119

Notes

1. In current usage, the term “ecological validity” has taken on a different meaning, being often used to describe the degree to which an experimental setting is a realistic representation of the real-world context to which the researcher hopes to generalize his or her research findings. Brunwik’s term for this kind of realism was “representativeness.”
2. Support for this project was awarded to Tera Letzring, Judith Hall, and Jeremy Biesanz by the National Science Foundation, award number 1551822.